

# STATE ESTIMATION FOR TEMPORAL POINT PROCESSES

M. N. M. van Lieshout

*CWI*

*P.O. Box 94079, NL-1090 GB Amsterdam, The Netherlands*

*Department of Applied Mathematics, University of Twente*

*P.O. Box 217, NL-7500 AE Enschede, The Netherlands*

This paper is concerned with combined inference for point processes on the real line observed in a broken interval. For such processes, the classic history-based approach cannot be used. Instead, we adapt tools from sequential spatial point processes. For a range of models, the marginal and conditional distributions are derived. We discuss likelihood based inference as well as parameter estimation using the method of moments, conduct a simulation study for the important special case of renewal processes and analyse a data set collected by Diggle and Hawtin.

*AMS Mathematics Subject Classification (2010):* 60G55; 60K15; 62M99.

*Key words & Phrases:* Cox process, Markov chain Monte Carlo, Markov point process, renewal process, sequential point process, state estimation.

## 1 Introduction

Inference for point processes on the real line has been dominated by a dynamic approach based on the stochastic intensity [4, 16, 17] which expresses the likelihood of a point at any given time conditional on the history of the process. Such an approach is quite natural in that it is the mathematical translation of the intuitive idea that more and more information becomes available as time passes. Furthermore, the approach allows the utilisation of powerful tools from martingale theory. In a statistical sense, since the stochastic intensity is closely related to the hazard rates of the distribution of the length of the inter-point intervals, a likelihood is immediately available [8].

The dynamic approach, however, does not seem capable of dealing with situations in which the flow of time is interrupted. In such cases, combined state estimation techniques are needed that are able to simultaneously carry out inference and to reconstruct the missing points. More specifically, state estimation aims to find ‘the optimal reconstruction, realization by realization, of unobserved portions of a point process or of associated random variables or processes’ [16, p. 93]. Other examples include extrapolation and prediction [24], filtering [28], cluster detection [19] or the estimation

of the driving measure of a Cox process [23].

The aim of this paper is to apply ideas from sequential point process [20, 21], in particular the sequential Papangelou conditional intensity which describes the probability of finding a point at a particular time conditional on the remainder of the process. Thus, the concept is related to the stochastic intensity, except that the future is taken into account as well as the past. It is this last feature that allows the incorporation of missing data. Moreover, for hereditary point processes at least, they define a likelihood.

The plan of this paper is as follows. In Section 2, we represent a point processes on the real line as a sequential point process on semi-cubes and recall the definition of the sequential Papangelou conditional intensity. Section 3 derives the marginal and conditional distributions for the situation where the point process is observed in disjoint intervals only. Important special cases are studied in detail: Markov point processes in Section 4.1, renewal processes in Section 4.2 and Cox processes in Section 5. In Section 6.1, we show that a substantial bias may be incurred when missing data is not accounted for properly. We present a Metropolis–Hastings algorithm for sampling from the conditional distribution of the missing data as well as a locally defined birth-and-death process based on the sequential Papangelou conditional intensity in Section 6.2 and the Appendix. These techniques are then used in Monte Carlo inference. Finally, in Section 7, a health care surveillance data set collected by Diggle and Hawtin is analysed to illustrate the approach and we conclude with a summary.

## 2 Finite point process on the line

In this section, we review basic concepts from the theory of chronologically ordered, simple point processes on an interval  $[0, T]$ ,  $T > 0$ .

Realisations of a simple point process  $X$  consist of a finite number of distinct points  $\{t_1, \dots, t_n\} \subset [0, T]$  for  $n \in \mathbb{N}_0$ . Since the points are naturally ordered, there is a unique correspondence between the set  $\{t_1, \dots, t_n\}$  and the vector  $(t_1, \dots, t_n)$  where  $t_1 < t_2 < \dots < t_n$  [8, 17]. The special case  $n = 0$  corresponds to an empty realisation. Equipping  $H_n([0, T]) = \{(t_1, \dots, t_n) \in [0, T]^n : t_1 < \dots < t_n\}$ ,  $n \in \mathbb{N}$ , with the product Borel  $\sigma$ -algebra, the distribution of a finite sequential point process  $Y$  can be specified by [20].

- a probability mass function  $q_n$ ,  $n \in \mathbb{N}_0$ , for the number of points in  $[0, T]$ ;
- for each  $n \in \mathbb{N}$  for which  $q_n > 0$ , a probability density function  $p_n$  on  $H_n([0, T])$  for the chronologically ordered vector of point locations given that there are  $n$  of them.

Note that the  $p_n$  need not be symmetric.

The likelihood of finding  $n \in \mathbb{N}$  points, one at each of the locations  $t_1, \dots, t_n$  in that order, is expressed by the Janossy densities [8]

$$j_n(t_1, \dots, t_n) = q_n p_n(t_1, \dots, t_n)$$

on  $H_n(T)$ . For  $n = 0$ ,  $j_0 = q_0$ . More compactly, one may specify a density  $f$  on  $N^f([0, T]) = \cup_{n=0}^{\infty} H_n([0, T])$  (equipped with the  $\sigma$ -algebra generated by the Borel product  $\sigma$ -fields) with respect to the probability measure

$$\nu(F) = \nu_{[0, T]}(F) = \sum_{n=0}^{\infty} e^{-T} \int_0^T \int_{t_1}^T \cdots \int_{t_{n-1}}^T 1_F(t_1, \dots, t_n) dt_1 \cdots dt_n$$

which is related to the unnormalised Janossy densities as  $f(t_1, \dots, t_n) = e^T j_n(t_1, \dots, t_n)$  and  $f(\emptyset) = e^T q_0$ . Note that  $\nu_{[0, T]}$  is the distribution of a unit rate Poisson process on  $[0, T]$ .

The sequential Papangelou conditional intensity [21] for inserting  $t$  at position  $k \in \{1, \dots, n+1\}$  is defined by

$$\lambda_k(t|t_1, \dots, t_n) = \frac{f(t_1, \dots, t_{k-1}, t, t_k, \dots, t_n)}{f(t_1, \dots, t_n)} \quad (1)$$

whenever both  $t_{k-1} < t < t_k$  and  $f(t_1, \dots, t_n) > 0$ ; it is set to zero otherwise. Reversely, provided  $f(\cdot)$  is hereditary in the sense that  $f(t_1, \dots, t_n) > 0$  implies that  $f(s_1, \dots, s_m) > 0$  for all subsequences  $(s_1, \dots, s_m)$  of  $(t_1, \dots, t_n)$ , the density factorises as

$$f(t_1, \dots, t_n) = f(\emptyset) \prod_{i=1}^n \lambda_i(t_i|t_1, \dots, t_{i-1}). \quad (2)$$

At this point, it is important to note that the term Papangelou conditional intensity is often abbreviated to plain ‘conditional intensity’. We choose not to do so to avoid confusion with another factorisation often referred to by this name. Indeed, by [8, Prop. 7.2.III], the probability density function of  $Y$  can be factorised as

$$f(t_1, \dots, t_n) = e^T \prod_{i=1}^n h^*(t_i) \exp \left[ - \int_0^T h^*(t) dt \right]$$

where  $h^*$  is a combination of the hazard functions of the conditional probability density functions  $\pi_n(t_n|t_1, \dots, t_{n-1})$  that govern the location of the  $n$ -th point conditional on the ‘past’ points  $t_1, \dots, t_{n-1}$ . We shall refer to  $h^*$  as the ‘hazard rate’. Note that one sometimes writes  $h^*(t_i) = h^*(t_i|t_1, \dots, t_{i-1})$  to stress the dependence on the history of  $Y$  up to but not including  $t_i$ . The two concepts are related, but not identical in general. For a rigorous treatment, the reader is referred to [16, 17].

### 3 Partially observed point processes

Consider a sequential point process that is observed in a broken interval  $[0, T_1] \cup [T_2, T]$ . In other words, any point falling in  $(T_1, T_2)$  is not observed. Thus, a realisation consists of two chronologically ordered sequences, say  $\vec{r} = (r_1, \dots, r_k)$ ,  $k \in \mathbb{N}_0$ , in  $[0, T_1]$  and  $\vec{s} = (s_1, \dots, s_l)$ ,  $l \in \mathbb{N}_0$ , in  $[T_2, T]$ .

In order to carry out likelihood based inference or state estimation, we need, respectively, the marginal distribution of the observations and the conditional law of the missing points.

**Proposition 1** *Let  $Y$  be a simple sequential point process on  $[0, T]$ ,  $T > 0$ , defined by  $p_n^Y$ ,  $q_n^Y$  as in Section 2.*

- (a) *The marginal distribution of the sequential point process  $Z$  defined as the restriction of  $Y$  to  $[0, T_1] \cup [T_2, T]$  has Janossy density*

$$j^Z(\vec{r}, \vec{s}) = \sum_{n=0}^{\infty} q_{n+n(\vec{r})+n(\vec{s})}^Y \int_{H_n((T_1, T_2))} p_{n+n(\vec{r})+n(\vec{s})}^Y(\vec{r}, t_1, \dots, t_n, \vec{s}) \prod_{i=1}^n dt_i \quad (3)$$

where  $n(\cdot)$  denotes vector length (with the convention that the integral for  $n = 0$  is equal to  $p_{n(\vec{r})+n(\vec{s})}^Y(\vec{r}, \vec{s})$ ). In other words,  $Z$  has density  $f^Z(\vec{r}, \vec{s}) = e^{T-T_2+T_1} j^Z(\vec{r}, \vec{s})$  with respect to  $\nu_{[0, T_1] \cup [T_2, T]}$ .

- (b) *If  $\vec{r}$  and  $\vec{s}$  are feasible in the sense that their marginal Janossy density (3) is strictly positive, then the conditional distribution of  $Y$  on  $(T_1, T_2)$  given  $\vec{r}$  and  $\vec{s}$  has Janossy density*

$$j_n(t_1, \dots, t_n | \vec{r}, \vec{s}) = \frac{1}{j^Z(\vec{r}, \vec{s})} q_{n+n(\vec{r})+n(\vec{s})}^Y p_{n+n(\vec{r})+n(\vec{s})}^Y(\vec{r}, t_1, \dots, t_n, \vec{s}) \quad (4)$$

for  $n \in \mathbb{N}$ ,  $q_0(\vec{r}, \vec{s}) = q_{n(\vec{r})+n(\vec{s})}^Y / j^Z(\vec{r}, \vec{s})$ . Equivalently, a conditional density with respect to  $\nu_{(T_1, T_2)}$  is  $f(t_1, \dots, t_n | \vec{r}, \vec{s}) = e^{T_2-T_1} j_n(t_1, \dots, t_n | \vec{r}, \vec{s})$ .

**Proof:** Split  $[0, T]$  in three parts:  $[0, T_1]$ ,  $(T_1, T_2)$  and  $[T_2, T]$ . Then,

$$N^f = \cup_{n=0}^{\infty} H_n([0, T]) = \cup_{n=0}^{\infty} \cup_{k,j,l:k+j+l=n}^{\infty} H_k([0, T_1]) \times H_j((T_1, T_2)) \times H_l([T_2, T]).$$

Write  $H_{k,j,l}(T_1, T_2, T) = H_k([0, T_1]) \times H_j((T_1, T_2)) \times H_l([T_2, T])$ . Then, the  $\sigma$ -algebra on the disjoint union set  $\cup_{j+k+l=n} H_{k,j,l}(T_1, T_2, T)$  can be described as follows. A subset  $B$  is measurable exactly if  $B \cap H_{k,j,l}(T_1, T_2, T)$  is a Borel set for all  $k, j, l$  adding up to  $n$  [12]. Clearly this holds for all Borel sets in  $H_n([0, T])$  and, reversely, any  $B$  measurable in the union  $\sigma$ -algebra is the disjoint union of Borel sets, hence a Borel set itself. Consequently,  $Y$  can also be split into three well-defined, though possibly dependent, sequential point processes by restriction to  $[0, T]$ ,  $(T_1, T_2)$  and  $[T_2, T]$ .

For  $k, j, l \in \mathbb{N}_0$ , write  $F_{k,j,l}$  for the event that  $j$  points fall in  $(T_1, T_2)$ ,  $k$  points in  $[0, T_1]$  and  $l$  further points in  $[T_2, T]$ . Then

$$\mathbb{P}(F_{k,j,l}) = q_{k+j+l}^Y \int_{H_k([0, T_1])} \int_{H_j((T_1, T_2))} \int_{H_l([T_2, T])} p_{k+j+l}^Y(\vec{\mathbf{r}}, \vec{\mathbf{t}}, \vec{\mathbf{s}}) d\vec{\mathbf{r}} d\vec{\mathbf{t}} d\vec{\mathbf{s}}$$

when  $k + j + l > 0$  and  $q_0^Y$  otherwise.

(a) For  $n \in \mathbb{N}_0$ , the marginal probability  $q_n$  of placing  $n$  points in  $[0, T_1] \cup [T_2, T]$  is the probability of the disjoint union  $\cup \{F_{k,j,l} : k, j, l \in \mathbb{N}_0, k + l = n\}$ , that is,

$$q_n^Z = \sum_{k, l \in \mathbb{N}_0 : k+l=n} \sum_{j=0}^{\infty} \mathbb{P}(F_{k,j,l}).$$

Similarly, suppose that  $q_n^Z > 0$  and condition on having  $n$  points in  $[0, T_1] \cup [T_2, T]$ . Then, for any measurable  $A \subseteq N^f([0, T_1] \cup [T_2, T])$ ,

$$\begin{aligned} \mathbb{P}(Z \in A | n(Z) = n) &= \sum_{k, l \in \mathbb{N}_0 : k+l=n} \sum_{j=0}^{\infty} \mathbb{P}(\{Z \in A\} \cap F_{k,j,l} | n(Z) = n) \\ &= \frac{1}{q_n^Z} \sum_{k, l \in \mathbb{N}_0 : k+l=n} \sum_{j=0}^{\infty} \mathbb{P}(\{Z \in A\} \cap F_{k,j,l}). \end{aligned}$$

Since  $\mathbb{P}(\{Z \in A\} \cap F_{k,j,l})$  can be written as

$$q_{k+j+l}^Y \int_{H_k([0, T_1])} \int_{H_j((T_1, T_2))} \int_{H_l([T_2, T])} 1_A(\vec{\mathbf{r}}, \vec{\mathbf{s}}) p_{k+j+l}^Y(\vec{\mathbf{r}}, \vec{\mathbf{t}}, \vec{\mathbf{s}}) d\vec{\mathbf{r}} d\vec{\mathbf{t}} d\vec{\mathbf{s}},$$

a particular configuration consisting of  $\vec{\mathbf{r}} \in H_k([0, T_1])$  and  $\vec{\mathbf{s}} \in H_l([T_2, T])$  with  $k + l = n$  has (conditional) probability density

$$\frac{1}{q_n^Z} \sum_{j=0}^{\infty} q_{n+j}^Y \int_{H_j((T_1, T_2))} p_{n+j}^Y(\vec{\mathbf{r}}, \vec{\mathbf{t}}, \vec{\mathbf{s}}) d\vec{\mathbf{t}}$$

with respect to Lebesgue measure on  $H_n([0, T_1] \cup [T_2, T])$ .

(b) Next, condition on  $\vec{\mathbf{r}} \in H_k([0, T_1])$  and  $\vec{\mathbf{s}} \in H_l([T_2, T])$ . To compute the probability of having  $n$ ,  $n \in \mathbb{N}_0$ , points in between  $T_1$  and  $T_2$ , we may restrict ourselves to  $F_{k,n,l}$ . If  $\mathbb{P}(F_{k,n,l}) = 0$ , this probability is zero. Thus, assume  $\mathbb{P}(F_{k,n,l}) > 0$  and, a fortiori,  $q_{k+n+l}^Y > 0$ . On  $F_{k,n,l}$ , the (joint) likelihood of the  $k + n + l$  points is  $q_{k+n+l}^Y p_{k+n+l}^Y(\cdot)$ , hence integration over  $H_n((T_1, T_2))$  yields that the marginal

density of  $(\vec{\mathbf{r}}, \vec{\mathbf{s}})$  with respect to Lebesgue measure on  $H_k([0, T_1]) \times H_l([T_2, T])$  is given by

$$q_{k+n+l}^Y \int_{H_n((T_1, T_2))} p_{k+n+l}^Y(\vec{\mathbf{r}}, \vec{\mathbf{t}}, \vec{\mathbf{s}}) d\vec{\mathbf{t}}.$$

We conclude that the conditional probability of finding  $n$  points in  $(T_1, T_2)$  given  $\vec{\mathbf{r}}$  and  $\vec{\mathbf{s}}$  elsewhere is

$$q_n(\vec{\mathbf{r}}, \vec{\mathbf{s}}) = \frac{q_{k+n+l}^Y \int_{H_n((T_1, T_2))} p_{k+n+l}^Y(\vec{\mathbf{r}}, \vec{\mathbf{t}}, \vec{\mathbf{s}}) d\vec{\mathbf{t}}}{\sum_{j=0}^{\infty} q_{k+j+l}^Y \int_{H_j((T_1, T_2))} p_{k+j+l}^Y(\vec{\mathbf{r}}, \vec{\mathbf{t}}, \vec{\mathbf{s}}) d\vec{\mathbf{t}}}.$$

The denominator is strictly positive by assumption. Moreover, a conditional density function for the locations of points in  $Y \cap (T_1, T_2)$  given there are  $n$  such points and given realisations  $\vec{\mathbf{r}}$  and  $\vec{\mathbf{s}}$  elsewhere exists and reads

$$\frac{p_{k+n+l}^Y(\vec{\mathbf{r}}, \dots, \vec{\mathbf{s}})}{\int_{H_n((T_1, T_2))} p_{k+n+l}^Y(\vec{\mathbf{r}}, \vec{\mathbf{t}}, \vec{\mathbf{s}}) d\vec{\mathbf{t}}}$$

provided the denominator is strictly positive, which is always the case if  $q_n(\vec{\mathbf{r}}, \vec{\mathbf{s}}) > 0$ .

□

For hereditary sequential point processes, we have the following simplification of Proposition 1(b).

**Proposition 2** *Let  $Y$  be a simple hereditary sequential point process with density  $f^Y$  with respect to  $\nu_{[0, T]}$ . Let  $\vec{\mathbf{r}} \in H_k([0, T_1])$  and  $\vec{\mathbf{s}} \in H_l([T_2, T])$  be feasible in the sense that  $j^Z(\vec{\mathbf{r}}, \vec{\mathbf{s}}) > 0$ . Then the conditional distribution of  $Y$  on  $(T_1, T_2)$ ,  $0 < T_1 < T_2 < T$ , given  $\vec{\mathbf{r}}$  and  $\vec{\mathbf{s}}$  is hereditary with Papangelou conditional intensity*

$$\lambda_{n+1}(t|t_1, \dots, t_n; \vec{\mathbf{r}}, \vec{\mathbf{s}}) = \lambda_{n+k+1}^Y(t|\vec{\mathbf{r}}, t_1, \dots, t_n, \vec{\mathbf{s}}) \quad (5)$$

*equal to the Papangelou conditional intensity of  $Y$  for all  $(t_1, \dots, t_n) \in H_n((T_1, T_2))$  and  $t \in (t_n, T_2]$  provided  $f^Y(\vec{\mathbf{r}}, t_1, \dots, t_n, \vec{\mathbf{s}}) > 0$ .*

**Proof:** By the proof of Proposition 1 part (b),

$$q_0(\vec{\mathbf{r}}, \vec{\mathbf{s}}) = \frac{q_{k+l}^Y p_{k+l}^Y(\vec{\mathbf{r}}, \vec{\mathbf{s}})}{j^Z(\vec{\mathbf{r}}, \vec{\mathbf{s}})}.$$

Therefore  $q_0(\vec{\mathbf{r}}, \vec{\mathbf{s}}) = 0$  would imply that  $j_{k+l}^Y(\vec{\mathbf{r}}, \vec{\mathbf{s}})$  and hence  $f^Y(\vec{\mathbf{r}}, \vec{\mathbf{s}})$  would be zero. Since  $Y$  is hereditary,  $f^Y(\vec{\mathbf{r}}, t_1, \dots, t_j, \vec{\mathbf{s}})$  would be zero for all  $t_1 < \dots < t_j$  and all

$j \in \mathbb{N}_0$ . Consequently,  $j^Z(\vec{r}, \vec{s})$  would be zero, a contradiction with the assumption. We conclude that  $q_0(\vec{r}, \vec{s}) > 0$ .

For  $n \in \mathbb{N}$ , consider the chronologically ordered vector  $(t_1, \dots, t_n)$  with components in  $(T_1, T_2)$  and assume that  $j_n(t_1, \dots, t_n | \vec{r}, \vec{s}) > 0$ . By (4) and the assumption that  $j^Z(\vec{r}, \vec{s}) > 0$ ,  $j_{k+n+l}^Y(\vec{r}, t_1, \dots, t_n, \vec{s}) > 0$ . Since  $Y$  is hereditary, all subsequences also have strictly positive Janossy density, in particular those including both  $\vec{r}$  and  $\vec{s}$ . Another appeal to (4) implies that the conditional point process is hereditary too.

By the general theory outlined in Section 2, the density factorises as

$$f(t_1, \dots, t_n | \vec{r}, \vec{s}) = e^{T_2 - T_1} q_0(\vec{r}, \vec{s}) \prod_{i=1}^n \lambda_i(t_i | t_1, \dots, t_{i-1}; \vec{r}, \vec{s})$$

in terms of the Papangelou conditional intensity

$$\lambda_i(t_i | t_1, \dots, t_{i-1}; \vec{r}, \vec{s}) = \frac{f(t_1, \dots, t_i | \vec{r}, \vec{s})}{f(t_1, \dots, t_{i-1} | \vec{r}, \vec{s})} = \frac{f^Y(\vec{r}, t_1, \dots, t_i, \vec{s})}{f^Y(\vec{r}, t_1, \dots, t_{i-1}, \vec{s})}$$

for  $f(t_1, \dots, t_{i-1} | \vec{r}, \vec{s}) > 0$  and zero otherwise. The last equality follows from Proposition 1 part (b) and implies (5).  $\square$

The Papangelou conditional intensity, in contrast to the hazard rate which is a function of the history of the process, is able to take into account past as well as future points. Indeed, there does not seem to be an analogue of (5) in terms of  $h^*$ .

The theorems above cannot be simplified without specific model assumptions. In the next sections we therefore specialise to, respectively, Markov point processes, nearest-neighbour processes and Cox models.

## 4 Markov point process

In this section, we consider models in which the dependence of a sequential point process in  $(T_1, T_2)$  on its future and past is limited either to a border region  $[T_1 - R, T_1] \cup [T_2, T_2 + R]$  for some fixed interaction range  $R$  (Section 4.1) or by its two nearest neighbours, say  $r_k$  in  $[0, T_1]$  and  $s_1 \in [T_2, T]$  (Section 4.2).

### 4.1 Fixed range dependence

A chronologically ordered sequential point process  $Y$  on  $[0, T]$  is said to be Markov at range  $R > 0$  if it is hereditary and and, for all  $(t_1, \dots, t_n)$  for which  $f^Y(t_1, \dots, t_n) > 0$  and for all  $u \neq t_i$ ,  $i = 1, \dots, n$ , the Papangelou conditional intensity  $\lambda_{n+1}^Y(u | t_1, \dots, t_n)$  depends only on  $u$  and the set of points  $t_i < u$  that have temporal distance  $u - t_i \leq R$ .

In other words, the underlying unordered point process is Markov in the classical sense [26]. Therefore, by the Hammersley–Clifford theorem, the density of  $Y$  factorises as

$$f^Y(t_1, \dots, t_n) = 1\{t_1 < \dots < t_n\} \prod \psi(\mathbf{y})$$

for some non-negative integrable interaction function  $\psi$ , ranging over the collection of *unordered* finite subsets  $\mathbf{y}$  of  $\{t_1, \dots, t_n\} \subseteq [0, T]$ , that vanishes except on cliques. More precisely,  $\psi(\mathbf{y}) = 1$  unless it consists of pairwise  $R$ -close points. To get rid of the indicator function, define  $\varphi(u, \mathbf{y}) = \psi(\mathbf{y} \cup \{u\})1\{u > \max(\mathbf{y})\}$  for  $u \in [0, T]$  so that

$$f^Y(t_1, \dots, t_n) = f^Y(\emptyset) \prod_{i=1}^n \prod_{\mathbf{t} \subseteq \{t_1, \dots, t_{i-1}\}} \varphi^Y(t_i, \mathbf{t})$$

cf. [20]. As usual, for  $n = 0$ ,  $(t_1, \dots, t_0)$  is interpreted as the empty sequence.

Next, turn to (5) and observe that it reduces to

$$\lambda_{n+1}(t|t_1, \dots, t_n; \vec{\mathbf{r}}, \vec{\mathbf{s}}) = \prod_{\mathbf{y} \subseteq \{t_1, \dots, t_n\}} \left[ \psi(\{t\} \cup \mathbf{y}) \prod_{\emptyset \neq \mathbf{x} \subseteq \mathbf{r} \cup \mathbf{s}} \psi(\{t\} \cup \mathbf{x} \cup \mathbf{y}) \right],$$

the product of all interaction functions that involve the point  $t$ . We conclude that the Markov property is preserved under conditioning but the clique interaction functions

$$\varphi(t, \mathbf{y}|\vec{\mathbf{r}}, \vec{\mathbf{s}}) = \varphi^Y(t, \mathbf{y}) \prod_{\emptyset \neq \mathbf{x} \subseteq \mathbf{r} \cup \mathbf{s}} \psi(\{t\} \cup \mathbf{x} \cup \mathbf{y})$$

for  $\mathbf{y} \subseteq \{t_1, \dots, t_n\}$  and  $t > t_n$  may change and depend on points of  $\vec{\mathbf{r}}$  and  $\vec{\mathbf{s}}$  in  $[T_1 - R, T_2 + R]$ . In particular, since  $\varphi(t, \emptyset|\vec{\mathbf{r}}, \vec{\mathbf{s}})$  need not be constant in  $t$  even when  $\varphi^Y(t, \mathbf{y})$  is, non-homogeneity may be introduced by the conditioning.

The special case of pairwise interaction processes of the form

$$f^Y(t_1, \dots, t_n) = f^Y(\emptyset) \prod_{i=1}^n \varphi^Y(t_i, \emptyset) \prod_{i < j: t_j - t_i \leq R} \varphi^Y(t_j, \{t_i\}) \quad (6)$$

with  $(t_1, \dots, t_n) \in H_n([0, T])$  is of special interest. For such models, the conditional distribution of  $Y$  on  $(T_1, T_2)$ ,  $0 < T_1 < T_2 < T$ , given  $\vec{\mathbf{r}}$  on  $[0, T_1]$  and  $\vec{\mathbf{s}}$  on  $[T_2, T]$  has first and second order interaction functions

$$\begin{aligned} \varphi(t, \emptyset|\vec{\mathbf{r}}, \vec{\mathbf{s}}) &= \varphi^Y(t, \emptyset) \prod_{x \in \vec{\mathbf{r}}: t-x \leq R} \varphi^Y(t, \{x\}) \prod_{x \in \vec{\mathbf{s}}: x-t \leq R} \varphi^Y(x, \{t\}); \\ \varphi(t_2, \{t_1\}|\vec{\mathbf{r}}, \vec{\mathbf{s}}) &= \varphi^Y(t_2, \{t_1\}). \end{aligned}$$

Therefore the conditional point process is also Markov at range  $R$  with pairwise interactions only. The second order interaction function is unaffected by the conditioning. The first-order interaction function depends on points of  $\vec{\mathbf{r}}$  and  $\vec{\mathbf{s}}$  up to a range  $R$  from  $(T_1, T_2)$ .



## 4.2 Nearest neighbour dependence

A renewal process on  $[0, T]$  is defined as follows [16, Chapter 8]. Starting at time 0, let  $U_i$  be a sequence of independent and identically distributed inter-arrival times with common probability distribution function  $F_\pi$ . We assume that  $F_\pi$  is non-defective and absolutely continuous with density  $\pi$ . Its hazard function is denoted by  $h_\pi$ . Set  $T_0 = 0$  and  $T_i = T_{i-1} + U_i$  for  $i \in \mathbb{N}$ . Then those  $T_i$ ,  $i \geq 1$  falling in  $(0, T]$  form a simple sequential point process  $Y$ . For simplicity, we assume that the process starts at time zero, but other starting distributions may also be accommodated

Due to the independence assumptions in the model, the hazard rate of  $Y$  is particularly appealing. Write  $V_t$  for the backwards recurrence time at  $t$ , that is, the difference between  $t$  and the last event falling before or at time  $t$ . Then,  $h^*(t) = h_\pi(V_{t-})$  [16], so that the density  $f^Y$  with respect to  $\nu_{[0, T]}$  is given by

$$\begin{aligned} f^Y(t_1, \dots, t_n) &= e^T \prod_{i=1}^n h^*(t_i) \exp \left[ - \int_0^T h^*(t) dt \right] \\ &= e^T (1 - F_\pi(T - t_n)) \prod_{i=1}^n \pi(t_i - t_{i-1}) \end{aligned} \quad (7)$$

for  $(t_1, \dots, t_n) \in H_n([0, T])$  [8, 16], under the conventions that an empty product is set to one and that  $t_0 = 0$ .

The sequential point process  $Y$  is not necessarily hereditary, for example when  $\pi$  has small bounded support. Thus, we shall assume that  $\pi > 0$ . In this case,  $f^Y$  can be factorised in terms of its Papangelou conditional intensity

$$\lambda_{n+1}^Y(t | t_1, \dots, t_n) = \pi(t - t_n) \frac{1 - F_\pi(T - t)}{1 - F_\pi(T - t_n)} \quad (8)$$

for  $t_n < t \leq T$ . It depends on the configuration  $t_1, \dots, t_n$  only through the point  $t_n$  that is closest to  $t$ , no matter how large the distance  $t - t_n$ . Such models are known as nearest-neighbour Markov point processes [3]. Their density satisfies a factorisation similar to (6) for fixed range Markov processes. Indeed, for  $n \in \mathbb{N}$ ,

$$f^Y(t_1, \dots, t_n) = f^Y(\emptyset) \prod_{i=1}^n \varphi^Y(t_i, \emptyset) \prod_{i=2}^n \varphi^Y(t_i, \{t_{i-1}\} | t_1, \dots, t_n)$$

and  $f^Y(\emptyset) = e^T (1 - F_\pi(T))$ . The interaction functions are

$$\varphi^Y(t, \emptyset) = \frac{\pi(t)(1 - F_\pi(T - t))}{1 - F_\pi(T)}$$

and

$$\varphi^Y(u, \{t\} | t_1, \dots, t_n) = \frac{\pi(u - t) [1 - F_\pi(T)]}{\pi(u) [1 - F_\pi(T - t)]}$$

for  $u \in \{t_1, \dots, t_n\}$  and  $t = \max\{t_i : t_i < u\}$  with  $\varphi^Y(t, \mathbf{y}|t_1, \dots, t_n) = 1$  otherwise. Note that, as the pairwise interaction function involves two consecutive neighbours, it depends on the whole configuration in contrast to the models discussed in the previous subsection.

Next, consider the conditional distribution of  $Y$  on  $(T_1, T_2)$  given  $\vec{\mathbf{r}}$  on  $[0, T_1]$  and  $\vec{\mathbf{s}}$  on  $[T_2, T]$ . Formula (5) specialises as follows:

$$\lambda_{n+1}(t|t_1, \dots, t_n; \vec{\mathbf{r}}, \vec{\mathbf{s}}) = \frac{\pi(t - t_n)\pi(s_1 - t)}{\pi(s_1 - t_n)} 1\{t_n < t < T_2\} \quad (9)$$

with the convention that for  $n = 0$ ,  $t_0 = \max(\vec{\mathbf{r}})$  if  $\vec{\mathbf{r}} \neq \emptyset$  and  $t_0 = 0$  otherwise. Compared to (8), the survival probability  $1 - F_\pi(T - t)$  is replaced by  $\pi(s_1 - t)$ . If  $\vec{\mathbf{s}} = \emptyset$ , (9) reads

$$\lambda_{n+1}(t|t_1, \dots, t_n; \vec{\mathbf{r}}, \emptyset) = \frac{\pi(t - t_n)(1 - F_\pi(T - t))}{1 - F_\pi(T - t_n)} 1\{t_n < t < T_2\}$$

as in (8), except for the fact that  $t$  is restricted to  $(T_1, T_2)$ . We conclude that the conditional distribution is a nearest-neighbour Markov point process on  $(T_1, T_2)$ . Assuming  $\vec{\mathbf{s}} \neq \emptyset$ , for  $n \in \mathbb{N}$ , its density can be written as

$$\begin{aligned} f(t_1, \dots, t_n | \vec{\mathbf{r}}, \vec{\mathbf{s}}) &= f(\emptyset | \vec{\mathbf{r}}, \vec{\mathbf{s}}) \frac{\pi(t_1 - \max(\vec{\mathbf{r}}))\pi(s_1 - t_n)}{\pi(s_1 - \max(\vec{\mathbf{r}}))} \prod_{i=2}^n \pi(t_i - t_{i-1}) \\ &= f(\emptyset | \vec{\mathbf{r}}, \vec{\mathbf{s}}) \prod_{i=1}^n \varphi(t_i, \emptyset | \vec{\mathbf{r}}, \vec{\mathbf{s}}) \prod_{i=2}^n \varphi(t_i, \{t_{i-1}\} | t_1, \dots, t_n; \vec{\mathbf{r}}, \vec{\mathbf{s}}) \end{aligned}$$

with

$$\varphi(t, \emptyset | \vec{\mathbf{r}}, \vec{\mathbf{s}}) = \frac{\pi(t - \max(\vec{\mathbf{r}}))\pi(s_1 - t)}{\pi(s_1 - \max(\vec{\mathbf{r}}))},$$

under the convention  $\max(\emptyset) = 0$ , and

$$\varphi(u, \{t\} | t_1, \dots, t_n; \vec{\mathbf{r}}, \vec{\mathbf{s}}) = \frac{\pi(u - t)\pi(s_1 - \max(\vec{\mathbf{r}}))}{\pi(u - \max(\vec{\mathbf{r}}))\pi(s_1 - t)}$$

for  $u \in \{t_1, \dots, t_n\}$  and  $t = \max\{t_i : t_i < u\}$  with  $\varphi^Y(t, \mathbf{y}|t_1, \dots, t_n) = 1$  otherwise. For the special case  $\vec{\mathbf{s}} = \emptyset$ , the ratio  $\pi(s_1 - t)/\pi(s_1 - \max(\vec{\mathbf{r}}))$  should be replaced by  $(1 - F_\pi(T - t))/(1 - F_\pi(T - \max(\vec{\mathbf{r}})))$ . In conclusion, the boundary conditions  $\vec{\mathbf{r}}$  and  $\vec{\mathbf{s}}$  affect the interaction functions only through the two nearest neighbours  $\max(\vec{\mathbf{r}})$  and  $\min(\vec{\mathbf{s}})$ , provided they exist.

## 5 Cox processes

A Cox process  $Y$  is a two-stage stochastic process. More specifically, let  $\Lambda(t)$  be a random field on  $[0, T]$  and define the random measure  $\Psi$  by

$$\Psi(A) = \int_A \Lambda(t) dt$$

for all Borel sets  $A \subseteq [0, T]$ . Then, conditionally on  $\Lambda = \lambda$ ,  $Y$  is a chronologically ordered Poisson process on  $[0, T]$  with intensity function  $\lambda$ . Clearly, conditions must be imposed on  $\Lambda$  so that the integral representation of  $\Psi$  is well-defined, for example that  $\Lambda$  almost surely has continuous realisations [1, 2]. The density and Janossy measures of  $Y$  are found by integrating out over the distribution of  $\Lambda$ , that is,

$$j_n^Y(t_1, \dots, t_n) = \mathbb{E} \left[ e^{-\Psi([0, T])} \prod_{i=1}^n \Lambda(t_i) \right] \quad (10)$$

for all  $t_1 < \dots < t_n$  [24].

Since the Papangelou conditional intensity typically is not known explicitly, we focus on Proposition 1. Firstly, upon plugging in formula (10),

$$f(t_1, \dots, t_n | \vec{\mathbf{r}}, \vec{\mathbf{s}}) = \frac{e^{T_2 - T_1}}{j^Z(\vec{\mathbf{r}}, \vec{\mathbf{s}})} \mathbb{E}_\Lambda \left[ e^{-\Psi([0, T])} \prod_{u \in (\vec{\mathbf{r}}, \vec{\mathbf{s}})} \Lambda(u) \prod_{i=1}^n \Lambda(t_i) \right]$$

with the marginal Janossy density of  $Z$ , the restriction of  $Y$  to  $([0, T_1] \cup [T_2, T])$ , given by

$$j^Z(\vec{\mathbf{r}}, \vec{\mathbf{s}}) = \mathbb{E}_\Lambda \left[ \exp[-\Psi([0, T_1] \cup [T_2, T])] \prod_{u \in (\vec{\mathbf{r}}, \vec{\mathbf{s}})} \Lambda(u) \right]$$

since conditional on  $\Lambda$ ,  $Z$  is a Poisson process. Finally, splitting  $\exp[-\Psi([0, T])] = \exp[-\Psi((T_1, T_2)) - \Psi([0, T_1] \cup [T_2, T])]$ , we conclude that, provided  $\vec{\mathbf{r}}$  and  $\vec{\mathbf{s}}$  are feasible in the sense that  $j^Z(\vec{\mathbf{r}}, \vec{\mathbf{s}}) > 0$ , the conditional distribution of  $Y$  on  $(T_1, T_2)$ ,  $0 < T_1 < T_2 < T$ , given  $\vec{\mathbf{r}}$  and  $\vec{\mathbf{s}}$  is a Cox process with its driving random field distributed as the

$$\prod_{u \in (\vec{\mathbf{r}}, \vec{\mathbf{s}})} \Lambda(u) \exp \left[ - \int_{[0, T] \setminus [T_1, T_2]} \Lambda(t) dt \right]$$

weighting of the distribution of  $\Lambda$ .

As a clarifying example, consider the so-called compound Poisson process on  $[0, T]$  for which  $\Lambda$  is constant over  $[0, T]$ , taking either of the values  $\lambda_1 \neq \lambda_2$  with equal

probability. Then, writing  $N$  for the sum of the cardinalities of  $\vec{\mathbf{r}}$  and  $\vec{\mathbf{s}}$  and  $\tilde{T}$  for the length of  $[0, T_1] \cup [T_2, T]$ ,

$$j_n(t_1, \dots, t_n | \vec{\mathbf{r}}, \vec{\mathbf{s}}) = \frac{\sum_i \lambda_i^{n+N} \exp \left[ -\lambda_i(\tilde{T}) - \lambda_i(T_2 - T_1) \right]}{\sum_i \lambda_i^N \exp \left[ -\lambda_i(\tilde{T}) \right]}.$$

In other words, the conditioning has the effect of changing the probability of  $\lambda_i$  from a half to

$$\frac{\lambda_i^N e^{-\lambda_i \tilde{T}}}{\lambda_1^N e^{-\lambda_1 \tilde{T}} + \lambda_2^N e^{-\lambda_2 \tilde{T}}}.$$

through  $N$ , the number of observed points. A further example will be studied in detail in Section 7.

## 6 Simulation study: renewal processes

In this section, we will consider parameter and state estimation for a well-known renewal process model. It is shown that naive approaches may result in bias. Instead, we shall adapt techniques for edge correction in spatial data to sequential point processes on the line.

Figure 1 shows a realisation of a temporal renewal process on  $[0, 4]$  observed within  $[0, 1] \cup [3, 4]$  with Erlang inter-arrival probability density

$$\pi(x) = \frac{\lambda^\alpha}{(\alpha - 1)!} x^{\alpha-1} e^{-\lambda x}, \quad x \geq 0,$$

for  $\lambda = 40$  and  $\alpha = 2$ .

For renewal processes observed in an unbroken interval  $[0, T]$ ,  $T > 0$ , [16, Chapter 8] surveys two approaches to estimate the parameters of the inter-arrival distribution  $\pi$ . The first method is to treat the fully observed inter-arrival times  $L_i$ ,  $i = 1, \dots, N(T)$ , as a random sample from  $\pi$  and apply maximum likelihood or the method of moments to obtain parameter estimates. The second approach is to use the explicit representation of the likelihood in terms of the hazard rate  $h^*$  and apply maximum likelihood estimation directly. Note that the first approach applies equally to observations on broken intervals, but the second one does not.

### 6.1 Inference based on fully observed intervals

Denote the lengths of the fully observed inter-arrival intervals by  $l_1, \dots, l_n$ . If these would constitute a valid random sample, for  $\alpha$  fixed, the model would be an exponential family with sufficient statistic  $\sum_i l_i$ . The maximum likelihood estimator would

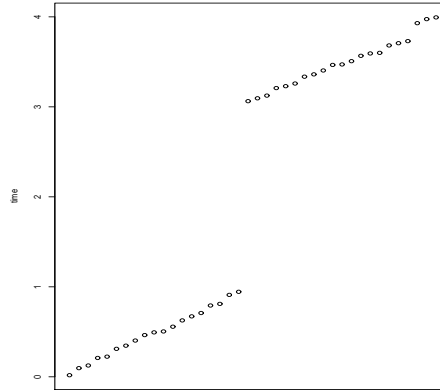


Figure 1: Sample from a renewal process with Erlang(2) inter-arrival times and rate parameter  $\lambda = 40$  observed in  $[0, 1] \cup [3, 4]$ . Time is plotted against index number.

exist and be given by  $\widehat{\lambda/\alpha} = n/\sum_i l_i$ . The parameter  $\alpha$  could be estimated by profile likelihood. In our context, however, the sample size  $n = N(T)$  is random. Nevertheless, [16] suggests to proceed as if the  $l_i$  were a random sample and claims this causes relatively little loss of information.

For the realisation consisting of 40 points depicted in Figure 1, consider the 38 observable inter-arrival times. For this sample,  $\hat{\lambda} = 40.91$ , whilst the profile likelihood method yields  $\hat{\alpha} = 2$ . To assess the bias and variance, we generated 100 data patterns on  $[0, 4]$  for the parameter values  $\lambda = 40$  and  $\alpha = 2$ . We iteratively excluded the middle of the left-most interval and estimated  $\lambda$ . The results are summarised in the table below.

observation interval	mean	variance
$[0, 1] \cup [3, 4]$	42.1	25.2
$[0, 0.25] \cup [0.75, 1]$	47.4	136.2
$[0, 0.0625] \cup [0.1875, 0.25]$	128.1	$1 \times 10^4$

We conclude that the bias and variance increase as the observation window contains less inter-arrival intervals. The bias occurs as smaller intervals are more likely to fall in the observation window, a phenomenon known as length bias [16]. For the smallest interval, the bias is severe. Indeed, of the samples on the union of two intervals of length  $1/16$  each, less than half had observable inter-arrival times at all. Moreover, the sub-interval lengths are only slightly larger than the expected inter-arrival length of  $1/20$ .

## 6.2 Monte Carlo maximum likelihood with missing data

The approach of Section 6.1 does not make full use of the available data, as it completely ignores partially observed inter-arrival intervals. Here, we adapt the method of [13] for dealing with missing data to better account for such intervals.

Again, suppose  $Y$  is a renewal process on  $[0, T]$  that is observed on  $[0, T_1] \cup [T_2, T]$ . Write  $\pi = \pi_\theta$  for the density of the inter-arrival time distribution. By part (b) of Theorem 1,  $j^Z(\vec{r}, \vec{s})$  is a normalising constant for the conditional Janossy density on  $(T_1, T_2)$  given the configuration elsewhere, say  $\vec{r}$  on  $[0, T_1]$  and  $\vec{s}$  on  $[T_2, T]$ . Hence, writing  $U$  for the vector of un-observed points in  $Y$  falling in  $(T_1, T_2)$ , the log likelihood ratio with respect to a given reference parameter  $\theta_0$  is [13, 24]

$$L(\theta) = \log j_\theta^Z(\vec{r}, \vec{s}) - \log j_{\theta_0}^Z(\vec{r}, \vec{s}) = \log \mathbb{E}_{\theta_0} \left[ \frac{f_\theta^Y(U, Z)}{f_{\theta_0}^Y(U, Z)} \mid Z = (\vec{r}, \vec{s}) \right],$$

where  $f_\theta^Y$  is of the form (7). In general, there is no explicit expression, so we must rely on Monte Carlo approximation, that is, replace the expectation by an average over a sample from the conditional distribution to obtain

$$L_N(\theta) = \log \left[ \frac{1}{N} \sum_{i=1}^N \frac{f_\theta^Y(\vec{r}, U_i^*, \vec{s})}{f_{\theta_0}^Y(\vec{r}, U_i^*, \vec{s})} \right]. \quad (11)$$

Specifically, the  $U_i^*$  form a sample from the conditional distribution of  $U$  on  $(T_1, T_2)$  given  $Z = (\vec{r}, \vec{s})$  on  $[0, T_1] \cup [T_2, T]$  under the reference parameter  $\theta_0$ . By taking derivatives, we get the Monte Carlo score and Fisher information. Note that even for exponential families, in the missing data case there may not be a unique maximum likelihood estimator.

To generate a sample  $U_i^*$ ,  $i = 1, \dots, N$ , we use the Metropolis–Hastings approach [6, 14, 15, 20] and tune it to the present context. An alternative is to use birth-and-death processes [25], but note that special care is needed to avoid explosion while ensuring sufficient mixing at the same time. Further details are provided in the Appendix.

Let us return to the model of Subsection 6.1 and the data plotted in Figure 1. The parameter is the rate  $\lambda$ , and, for  $0 < t_1 < \dots < t_n < T = 4$ ,

$$\frac{f_\lambda^Y(t_1, \dots, t_n)}{f_{\lambda_0}^Y(t_1, \dots, t_n)} = \left( \frac{\lambda}{\lambda_0} \right)^{n\alpha} \frac{e^{-\lambda T} \sum_{i=0}^{\alpha-1} \lambda^i (T - t_n)^i / i!}{e^{-\lambda_0 T} \sum_{i=0}^{\alpha-1} \lambda_0^i (T - t_n)^i / i!}.$$

Hence, the sufficient statistic consists of  $n$ , the number of points, and  $T - t_n$ , the backward recurrence time from  $T$ . Write  $\vec{r} = (r_1, \dots, r_k)$  for the points observed in  $[0, 1]$  and  $\vec{s} = (s_1, \dots, s_l)$  for those in  $[3, 4]$ . For our data,  $\vec{r}$  consists of 19 points,  $\vec{s}$  of 21.

To find a suitable reference parameter for the log likelihood ratio, we use the Newton–Raphson method [13]. For the data of Figure 1, this method gives  $\lambda_0 = 40.531$ .

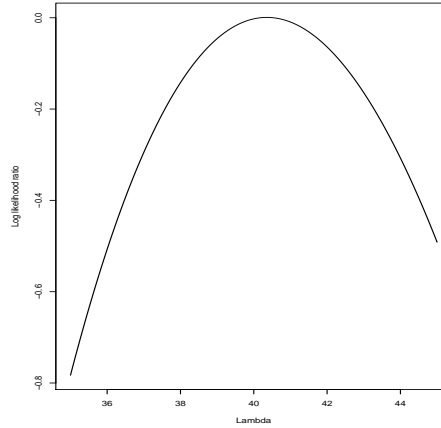


Figure 2: Monte Carlo log likelihood ratio for the data shown in Figure 1 based on 10,000 samples from a long run of the Metropolis–Hastings algorithm sub-sampled every thousand steps with  $\lambda_0 = 40.531$ .

We compute the Monte Carlo log likelihood ratio  $L_N(\lambda)$  using  $N = 10,000$  samples from a long Metropolis–Hastings run sub-sampled every thousand steps after a thousand steps burn-in. Note that it is sufficient to store the sufficient statistics only, in this case the length of the vectors  $U_i^*$ . The function is shown in Figure 2. It has a unique maximum at  $\hat{\lambda} = 40.36$ ; the Monte Carlo inverse Fisher information is 20.3.

To assess the bias and variance, we considered a hundred data patterns as in Section 6.1. For each pattern, after ten Newton–Raphson steps from the true value  $\lambda_0 = 40.0$ ,  $N = 1,000$  Monte Carlo samples were obtained by sub-sampling in a long run every thousand steps after a burn-in of a thousand steps. The results are summarised in the table below.

observation interval	mean ( $\lambda$ )	variance
$[0, 1] \cup [3, 4]$	40.9	22.8
$[0, 0.25] \cup [0.75, 1]$	40.8	76.7
$[0, 0.0625] \cup [0.1875, 0.25]$	40.8	457.3

One sees that compared to the naive approach, the bias is much reduced by correctly taking into account missing data and does not increase noticeably for smaller sub-intervals. The variance is also reduced but does increase when more data is missing.

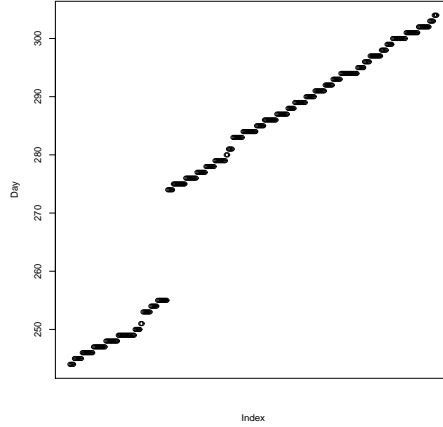


Figure 3: Calls to NHS Direct from Hampshire in September–October 2001. The date of calls, in days starting January 1st, 2001, is plotted against index number.

## 7 Application

Figure 3 shows daily records of calls to NHS Direct, a phone service operated by the National Health Service in Britain which people could call 24 hours a day to get medical advice. The calls shown are those that reported acute gastroenteric complaints in the county of Hampshire (on the English south coast and including e.g. the towns of Winchester and Southampton). The full data were described and analysed in [9, 10, 11]. A modified version provided by Professor Diggle and Dr Hawtin is available at

[www.lancaster.ac.uk/staff/diggle/pointpatternbook/datasets/AEGISS](http://www.lancaster.ac.uk/staff/diggle/pointpatternbook/datasets/AEGISS).

The latter data contain 7167 records from the years 2001 and 2002. Figure 3 plots the 327 calls recorded during September and October 2001. A salient feature is that no calls were registered during the period September 13th–September 30th with recording resuming on October 1st, 2001. This gap is clearly visible in the figure.

As in [9], we assume that calls are made according to a log-Gaussian Cox process on  $[0, T]$  with  $T = 61$  (time measured in days) so that in the notation of Section 5  $\Lambda(t) = \mu_0(t)e^{S(t)}$ . The integrable function  $\mu_0 : [0, T] \rightarrow \mathbb{R}^+$  accounts for temporal variations and  $S$  is an Ornstein–Uhlenbeck process, that is, a stationary, Gaussian, Markovian process with exponentially decaying covariance function

$$\text{Cov}(S(t), S(t')) = \sigma^2 e^{-\beta|t-t'|}.$$



The mean is chosen so that the expectation of  $\exp(S(t)) \equiv 1$ . By [1, 2, 23],  $S$  has almost surely continuous realisations and the model is well-defined.

Since the data is discretised in daily counts, the intensity of counts during the  $i$ -th day may be approximated by  $\mu_0(i)e^{S(i)}$ ,  $i = 0, \dots, 60$ . As  $S$  is an Ornstein–Uhlenbeck process, it satisfies the Markov recursion

$$\begin{aligned} S(0) &= -\frac{\sigma^2}{2} + \sigma\Gamma(0) \\ S(i) &= -\frac{\sigma^2}{2} (1 - e^{-\beta}) + e^{-\beta}S(i-1) + \sigma\Gamma(i) \end{aligned}$$

for independent zero-mean normally distributed random variables  $\Gamma(i)$  having variance  $1 - \exp(-2\beta)$  under the convention 1 for  $\Gamma(0)$ . In terms of the  $\Gamma(i)$ ,

$$S(i) = -\frac{\sigma^2}{2} + \sigma \sum_{j=0}^i e^{-\beta(i-j)} \Gamma(j). \quad (12)$$

In words,  $S(i)$  is a sum of independent normally distributed components discounted by elapsed time.

The goal of this section is two-fold. Firstly, we shall contrast moment based estimation with and without taking into account the gap; then, we will consider state estimation.

## 7.1 Parameter estimation

Although in principle it is possible to carry out likelihood based inference, this would be computationally costly since the missing data consist not only of the missing counts in September 2001, but also of *all* the  $S(i)$  or, equivalently, the  $\Gamma(i)$ , cf. Section 5. Hence [9] proposed to use moment methods.

First consider  $\mu_0$ . Since the expectation of  $e^{S(i)}$  is one,  $\mu_0(i)$  is the expected number of counts for day  $i$ . This observation led [9] to fit a Poisson log-linear regression model of the form

$$\log \mu_0(i) = \delta_{d(i)} + a_1 \cos\left(\frac{2i\pi}{365}\right) + b_1 \sin\left(\frac{2i\pi}{365}\right) + a_2 \cos\left(\frac{4i\pi}{365}\right) + b_2 \sin\left(\frac{4i\pi}{365}\right) + gi,$$

$i = 0, \dots, 60$ . This model takes into account the apparent increase of incidences in Spring, the overall increase in calls to NHS Direct over time and the weekend-effect through the  $d(i)$ , but ignores over-dispersion due to the Gaussian field. Note also that the counts are conditionally rather than jointly independent.

To assess the effect of the gap, we estimate the parameters based on two scenarios as follows. A naive estimator treats the zero counts in September as if they were true observations. The result is the red line in Figure 4. A more sophisticated approach

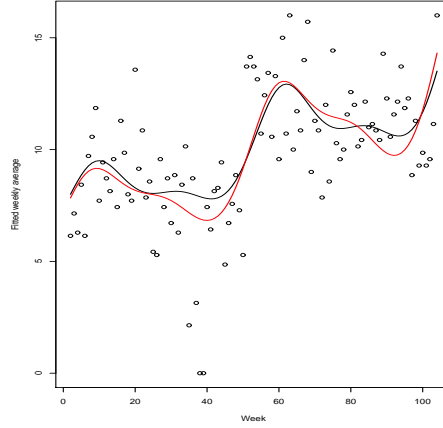


Figure 4: Weekly averages of calls to NHS Direct during 2001–20002 plotted against the week index. Red line: fit by naive approach; Black: fit by approach taking into account the gap.

disregards the time period from September 13–30, with the black line in Figure 4 as a result. The trough due to the missing data is clearly visible in the picture and, due to the periodic nature of the fitted model, also affects the behaviour a year later.

Next, turn to estimation of  $\sigma^2$  and  $\beta$ . Let  $N_i$ ,  $i = 0, \dots, I = 729$ , denote the observed number of calls during day  $i$ . The analogue of the pair correlation function [7] for count data is  $(N_i N_{i-\nu}) / (\mu_0(i) \mu_0(i - \nu))$  which has expectation  $\exp(\sigma^2 e^{-\beta|\nu|})$  at lag  $\nu \neq 0$  regardless of  $i$ . Therefore, the minimum contrast method minimises

$$\sum_{\nu=1}^m \left[ \frac{1}{I - \nu + 1} \sum_{i=\nu}^I \frac{N_i N_{i-\nu}}{\hat{\mu}_0(i) \hat{\mu}_0(i - \nu)} - \exp(\sigma^2 e^{-\beta\nu}) \right]^2 \quad (13)$$

over  $\beta$  and  $\sigma^2$ . Here,  $\hat{\mu}_0$  is an estimate of  $\mu_0$ ,  $I = 729$  and  $m$  is the number of lags considered.

Minimising (13) with  $\hat{\mu}_0$  based on all data including the spurious zeroes using  $m = 14$  lags, we obtain  $\hat{\sigma}^2 = 0.12$  and  $\hat{\beta} = 0.55$ . The fit is indicated in the left-most panel of Figure 5. Note that  $\mu_0$  is underestimated, hence the pair correlation function is overestimated.

The zero counts between September 13th and September 30th, 2001, should not be taken into account. Thus, in (13), we plug in the appropriate estimator  $\hat{\mu}_0$  (on which the black line in Figure 4 is based) and for lag  $\nu$ , consider only pairs  $(N_i, N_{i-\nu})$  for which both  $i$  and  $i - \nu$  do not fall in the period September 13–30. Doing so, we obtain  $\hat{\sigma}^2 = 0.11$  and  $\hat{\beta} = 0.91$ . Thus, the value of  $\sigma^2$  is not much affected, but that of  $\hat{\beta}$  is.

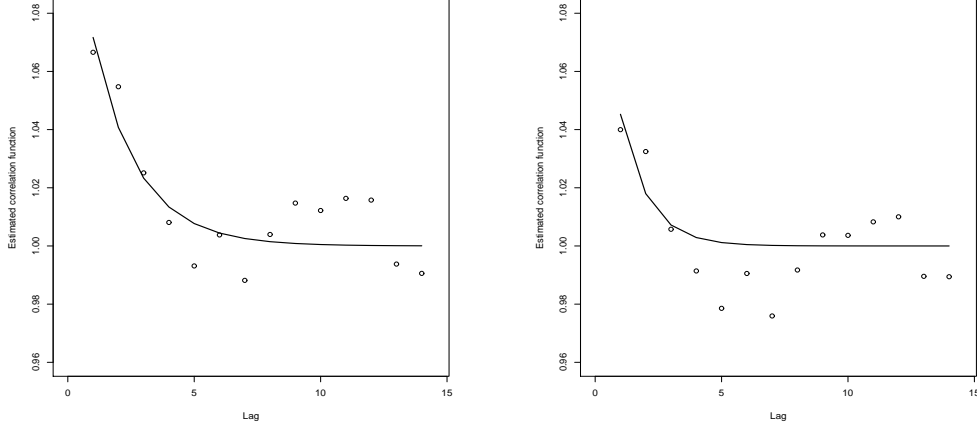


Figure 5: Minimum contrast fit for the naive approach (left) and when taking into account the missing data (right). The points are the mean of  $N_i N_{i-\nu} / (\hat{\mu}_0(i) \hat{\mu}_0(i - \nu))$  plotted as a function of the lag  $\nu$ .

The fit is indicated in the right-most panel of Figure 5. We conclude that correlations between counts separated by a week or more are small.

## 7.2 State estimation

Our final goal is to fill the gap in Figure 3 by sampling from the conditional distribution given the data. Now, as the estimated pair correlation function shown in Figure 5 is close to one for lags of a week and more, one may restrict attention to the counts  $n(T_i - 6), \dots, n(T_1)$  and  $n(T_2), \dots, n(T_2 + 6)$  either side of the gap. Moreover, for Cox processes, state estimation amounts to sampling from the driving random measure, cf. Section 5, which, by (12), is uniquely defined by the  $\Gamma(i)$ ,  $i = 0, \dots, T$ . Therefore, the log likelihood, up to a constant  $c(\beta, \sigma^2, \mu_0, n(i)_i)$ , is given by

$$-\frac{1}{2}\gamma(0)^2 - \sum_{i=1}^T \frac{\gamma(i)^2}{2(1 - e^{-2\beta})} + \sum_{i \in \{T_1-6, \dots, T_1, T_2, \dots, T_2+6\}} [n(i)S_\gamma(i) - \mu_0(i)e^{S_\gamma(i)}]. \quad (14)$$

Since  $c(\beta, \sigma^2, \mu_0, n(i)_i)$  cannot be evaluated exactly, we use a Metropolis–Hastings algorithm to sample from (14) as in [5]. The proposal distribution is a multivariate normal one with independent components having variance  $h > 0$  and mean  $h/2$  times

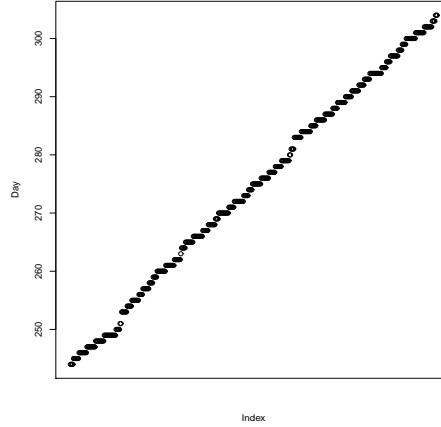


Figure 6: State estimation for the data shown in Figure 3. The date of calls is plotted against index number.

the gradient of (14). Thus, transitions are steered into a desirable direction. It is shown in [27] that the Markov chain is aperiodic and Lebesgue-irreducible, so that for almost all starting states, the chain converges to (14) in total variation.

For  $h = 0.5$ , we ran the chain for 10,000 steps. Using the realisation thus obtained, shown in Figure 6, we sampled the counts and re-estimated the parameters by minimum contrast. The results are similar to those obtained using the more sophisticated approach in Section 7.1. Indeed,  $\hat{\sigma}^2 = 0.11$  and  $\hat{\beta} = 0.93$ , whereas the plot of  $\hat{\mu}_0$  averaged over weeks is almost identical to the black line in Figure 4.

Next, we extended the run, sub-sampling every 1,000 steps, to obtain 10,000 realisations of (14) and calculated the histogram of the total intensity in the gap period. It is shown in Figure 7. In conclusion, around 150 calls may have been missed.

## 8 Summary

In this paper, we considered temporal point processes observed in broken observation windows. We derived the marginal and conditional distributions in terms of the Janossy densities of the underlying sequential point process and studied Markovian and Cox model in more detail. We carried out a simulation study to assess the length bias in renewal processes and analysed a real-life data set. The approach can easily be

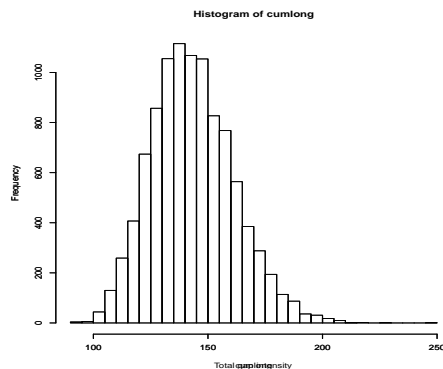


Figure 7: Histogram of the conditional total intensity of calls during September 13–30, 2001, given the observed calls plotted in Figure 3.

extended to space-time by incorporating spatial marks.

## Acknowledgements

Data were provided by Professor Peter Diggle, Lancaster University, and Dr Peter Hawtin, Health Protection Agency, Southampton Laboratory. The data were derived from anonymised data-sets collected during project AEGISS, a collaborative surveillance project based in the south of England. The project was financially supported by the Food Standards Agency. The author acknowledges Professors Boucherie and Van der Mei for fruitful discussions.

## References

- [1] R.J. Adler. *The geometry of random fields*. Chichester: Wiley, 1981.
- [2] R.J. Adler and J.E. Taylor. *Random fields and geometry*. New York: Springer, 2007.
- [3] A.J. Baddeley and J. Møller. Nearest-neighbour Markov point processes and random sets. *Internat. Statist. Rev.*, 57:89–121, 1989.
- [4] P. Brémaud. *A martingale approach to point processes*. Ph.D. dissertation, University of Berkeley, 1972.

- [5] A. Brix and P.J. Diggle. Spatiotemporal prediction for log-Gaussian Cox processes. *J. R. Statist. Soc. B*, 63:823–841, 2001.
- [6] S. Brooks, A. Gelman, G. Jones and X.-L. Meng (Eds). *Handbook of Markov Chain Monte Carlo*. Handbooks of Modern Statistical Methods. Boca Raton: Chapman & Hall/CRC Press, 2011.
- [7] S.N. Chiu, D. Stoyan, W.S. Kendall and J. Mecke. *Stochastic geometry and its applications*. 3rd ed. Chichester: Wiley, 2013.
- [8] D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Volume I: Elementary theory and methods*. 2nd ed. New York: Springer, 2003.
- [9] P.J. Diggle, B. Rowlingson and T.-L. Su. Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics*, 16:423–434, 2005.
- [10] P.J. Diggle, P. Moraga, B. Rowlingson and B.M. Taylor. Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Stat. Sci.*, 28:542–563, 2013.
- [11] P. Diggle, L. Knorr-Held, B. Rowlingson, T. Su, P. Hawtin and T. Bryant. *On-line monitoring of public health surveillance data*. In *Monitoring the health of populations: Statistical principles and methods for public health surveillance*, pp. 233–266. R. Brookmeyer and D.F. Stroup (Eds.). Oxford: Oxford University Press, 2003.
- [12] D.H. Fremlin. *Measure theory. Volume 2: Broad foundations*. 2nd ed. Colchester: Torres Fremlin, 2003.
- [13] C. Geyer. Likelihood inference for spatial point processes. In *Stochastic geometry (Toulouse, 1996)*. Monographs in Statistics and Applied Probability, Volume 80, pp. 79–140. Boca Raton: Chapman & Hall/CRC Press, 1999.
- [14] C.J. Geyer and J. Møller. Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Stat.*, 21:359–373, 1994.
- [15] P.J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82: 711–732, 1995.
- [16] A.F. Karr. *Point processes and their statistical inference*. 2nd ed. New York: Marcel Dekker, 1991.
- [17] G. Last and A. Brandt. *Marked point processes on the real line. The dynamic approach*. New York: Springer, 1995.

- [18] M.N.M. van Lieshout. *Markov point processes and their applications*. London/Singapore: Imperial College Press/World Scientific Publishing, 2000.
- [19] M.N.M. van Lieshout and A.J. Baddeley. Extrapolating and interpolating spatial pattern. *Spatial cluster modelling*, A.B. Lawson and D.G.T. Denison (Eds.), pp. 61–86. Boca Raton: Chapman & Hall/CRC Press, 2002.
- [20] M.N.M. van Lieshout. Markovianity in space and time. In *Dynamics & stochastics: Festschrift in honour of M.S. Keane*, D. Denteneer, F. den Hollander and E. Verbitskiy (Eds.), Lecture Notes–Monograph Series, Volume 48, pp. 154–168. Beachwood: Institute for Mathematical Statistics, 2006.
- [21] M.N.M. van Lieshout. Campbell and moment measures for finite sequential spatial processes. In *Proceedings Prague Stochastics 2006*, M. Hušková and M. Janžura (Eds.), pp. 215–224. Prague: Matfyzpress, 2006.
- [22] S.P. Meyn and T.L. Tweedie. *Markov chains and stochastic stability*. 2nd ed. Cambridge: Cambridge University Press, 2009.
- [23] J. Møller, A.R. Syversveen and R.P. Waagepetersen. Log Gaussian Cox processes. *Scand. J. Statist.*, 25:451–482, 1998.
- [24] J. Møller and R.P. Waagepetersen. *Statistical inference and simulation for spatial point processes*. Boca Raton: Chapman & Hall/CRC, 2004.
- [25] C.J. Preston. Spatial birth-and-death processes. *Bull. Inst. Internat. Statist.*, 46:371–391, 1977.
- [26] B.D. Ripley and F.P. Kelly. Markov point processes. *J. London Math. Soc. (2)*, 15:188–192, 1977.
- [27] G.O. Roberts and R.L. Tweedie. Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli*, 2:341–363, 1997.
- [28] S.S. Singh, B.-N. Vo, A. Baddeley, S. Zueyv. Filters for spatial point processes. *SIAM J. Control Optim.*, 48:2275–2295, 2009.

## Appendix: Markov chain Monte Carlo

To simulate a point process whose distribution is defined by the Papangelou conditional intensity, we adapt two general strategies to our context. Throughout, let  $Y$  be an hereditary sequential point process on  $[0, T]$  with density  $f$  with respect to  $\nu_{[0, T]}$  and Papangelou conditional intensity  $\lambda_i(\cdot|\cdot)$ .

### Metropolis–Hastings sampling

The Metropolis–Hastings method works by proposing an update according to a distribution that is convenient to sample from, and then to accept or reject this proposal with a probability that is chosen so as to make sure the detailed balance equations are satisfied [6].

The two generic types of proposals are births and deaths. More precisely, with probability  $1/2$  propose a birth, otherwise a death. In the first case, select a point  $u$  uniformly on  $[0, T]$  and insert it in its chronological position  $i$ . In case of a death, a point is chosen uniformly for deletion; if the sequence is empty, nothing happens. Then, if the current vector is  $(t_1, \dots, t_n)$ , the proposal to add  $u$  is accepted with probability

$$\min \left\{ 1, \frac{\lambda_i(u|t_1, \dots, t_n)T}{n+1} \right\};$$

the proposal to delete the  $i$ -th point with probability

$$\min \left\{ 1, \frac{n}{\lambda_i(t_i|t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n)T} \right\}.$$

It is easily seen that  $f(\cdot)$  is an invariant density. Moreover, if we start the chain in a sequence  $\vec{t}$  for which  $f(\vec{t}) > 0$ , the chain will almost surely never leave the set of states having positive density.

In order to show that the Metropolis–Hastings chain  $Y_n$ ,  $n \in \mathbb{N}_0$ , converges to  $f(\cdot)$  in total variation from any initial state having positive density, it is sufficient to assume stability, that is, that the Papangelou conditional intensity is uniformly bounded by some  $\beta > 0$ . The proof follows the same lines as in [13].

### Birth-and-death process sampling

Since the Papangelou conditional intensity may fluctuate a lot, we define birth-and-death processes in terms of local bounds. More precisely, assume that

$$\lambda_i(u|t_1, \dots, t_n) \leq g(u|t_1, \dots, t_n) \leq \beta \tag{15}$$

for some integrable function  $g$  that is constant on  $(t_{i-1}, t_i)$ ,  $i = 1, \dots, n+1$ , say  $g_i(t_1, \dots, t_n)$ , and some  $\beta > 0$ , with an appropriate convention for  $i = 1$  and  $i =$



$n + 1$ . Then, the birth-and-death algorithm runs as follows. If the current state is  $\vec{\mathbf{t}} = (t_1, \dots, t_n)$ ,

- compute the upper bound  $G(t_1, \dots, t_n)$  to the total birth rate by

$$G(t_1, \dots, t_n) = \sum_{i=1}^{n+1} \int_{t_{i-1}}^{t_i} g(u|t_1, \dots, t_n) du = \sum_{i=1}^{n+1} (t_i - t_{i-1}) g_i(t_1, \dots, t_n)$$

and take death rate  $D(t_1, \dots, t_n) = n$ ;

- generate an exponentially distributed sojourn time with rate parameter  $G(\vec{\mathbf{t}}) + D(\vec{\mathbf{t}})$ ;
- with probability  $\frac{G(\vec{\mathbf{t}})}{G(\vec{\mathbf{t}}) + D(\vec{\mathbf{t}})}$ , generate a new point ('birth') as follows:
  - sample an interval with probability  $\frac{(t_i - t_{i-1})g_i(\vec{\mathbf{t}})}{G(\vec{\mathbf{t}})}$  and propose a new point  $u$  uniformly in the chosen interval;
  - accept the proposal with probability  $\frac{\lambda_i(u|\vec{\mathbf{t}})}{g_i(\vec{\mathbf{t}})}$ ;
- with probability  $\frac{D(\vec{\mathbf{t}})}{G(\vec{\mathbf{t}}) + D(\vec{\mathbf{t}})}$ , delete a uniformly chosen point.

Then the rate for a transition from  $\vec{\mathbf{t}}$  to  $(t_1, \dots, t_i, u, t_{i+1}, \dots, t_n)$  for  $t_i < u < t_{i+1}$  is

$$(G(\vec{\mathbf{t}}) + D(\vec{\mathbf{t}})) \frac{G(\vec{\mathbf{t}})}{G(\vec{\mathbf{t}}) + D(\vec{\mathbf{t}})} \frac{(t_i - t_{i-1})g_i(\vec{\mathbf{t}})}{G(\vec{\mathbf{t}})} \frac{1}{t_i - t_{i-1}} \frac{\lambda_i(u|\vec{\mathbf{t}})}{g_i(\vec{\mathbf{t}})} = \lambda_i(u|\vec{\mathbf{t}})$$

so that the birth rate is effectively  $\lambda_i(u|t_1, \dots, t_n)$  on  $(t_{i-1}, t_i)$ . Therefore,  $f$  is an invariant density. An appeal to [25] implies the existence of a unique jump process with the given birth and death rates;  $f$  is its unique invariant probability density and the process converges to  $f$  in distribution from any  $\vec{\mathbf{t}}$  for which  $f(\vec{\mathbf{t}}) > 0$ .

**Renewal process** As an example, consider a renewal process with Erlang inter-arrival times. To derive a local bound on the conditional intensity, note that for  $a < \xi < b$ ,

$$\frac{\pi(\xi - a)\pi(b - \xi)}{\pi(b - a)} \leq \frac{\lambda^\alpha}{(\alpha - 1)!} \left( \frac{b - a}{4} \right)^{\alpha-1}$$

and

$$\frac{\pi(\xi - a)(1 - F_\pi(b - \xi))}{1 - F_\pi(b - a)} = \frac{\lambda^\alpha}{(\alpha - 1)!} (\xi - a)^{\alpha-1} \frac{\sum_{i=0}^{\alpha-1} \lambda^i (b - \xi)^i / i!}{\sum_{i=0}^{\alpha-1} \lambda^i (b - a)^i / i!}$$

is bounded from above by

$$\frac{\lambda^\alpha (b-a)^{\alpha-1}}{(\alpha-1)! \sum_{i=0}^{\alpha-1} \lambda^i (b-a)^i / i!} \sum_{i=0}^{\alpha-1} \left[ \frac{\lambda(b-a)}{4} \right]^i / i!.$$

The last bound follows from the fact that

$$\begin{aligned} (\xi - a)^{\alpha-1} (b - \xi)^i &= (\xi - a)^{\alpha-1-i} [(b - \xi)(\xi - a)]^i \\ &\leq (b - a)^{\alpha-1-i} \left[ \frac{(b - a)^2}{4} \right]^i = (b - a)^{\alpha-1+i} 4^{-i}. \end{aligned}$$

The bound can be improved upon in some cases by noting that the function

$$\phi : \xi \rightarrow \pi(\xi - a)(1 - F_\pi(b - \xi))$$

increases on  $(a, b)$  if  $\lambda - (\alpha - 1)/(b - a) < 0$ . Finally, since the sub-interval length  $b - a \leq T$  is bounded, stability follows.